

AN OVERVIEW OF THE EOS DATA DISSEMINATION SYSTEMS

H.K. Ramapriyan, Robin Pfister and Beth Weinstein

NASA Goddard Space Flight Center*

INTRODUCTION

The Earth Observing System Data and Information System (EOSDIS) is the primary data system serving the broad-scope of NASA's Earth Observing System (EOS) program and a significant portion of the "heritage" Earth science data. EOSDIS was designed to support the Earth sciences within NASA's Science Mission Directorate (previously the Earth Science Enterprise (ESE) and Mission to Planet Earth). The EOS Program was NASA's contribution to the United States Global Change Research Program (USGCRP) enacted by Congress in 1990 as part of the Global Change Act. ESE's objective was to launch a series of missions to help answer fundamental global change questions such as "How is Earth changing?" and "What are the consequences for life on Earth?" resulting support of this objective, EOSDIS distributes a wide variety of data (Figure DD1) to a diverse community.

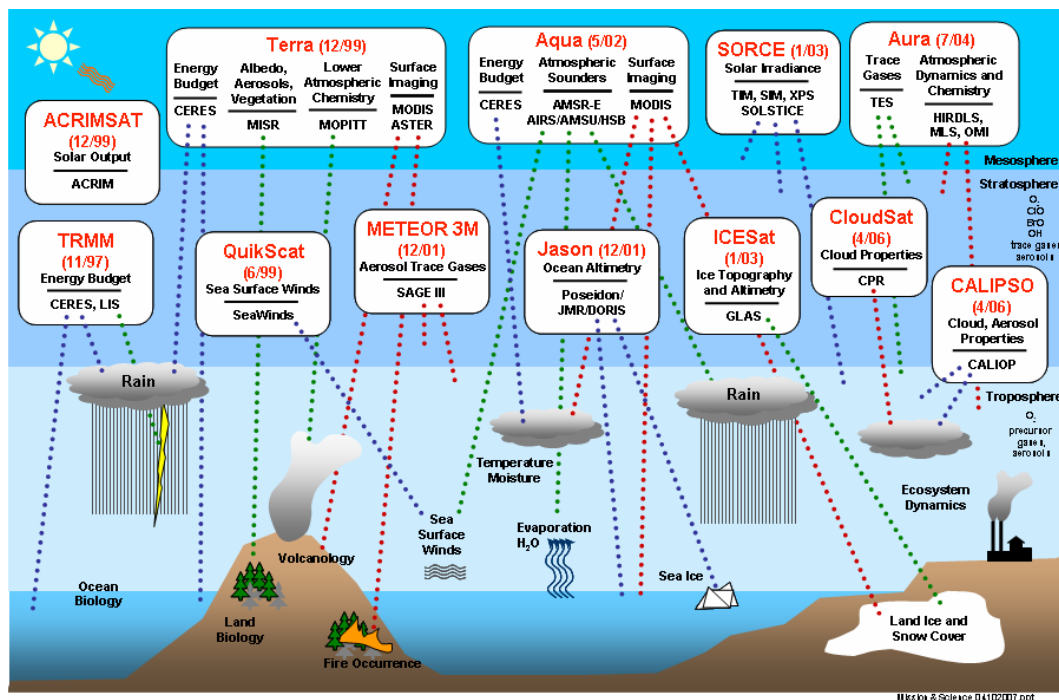


Figure DD1. EOS Missions, Instruments and Platforms

The EOSDIS data are held in 11 geographically distributed Data Centers. These centers are responsible for archiving and distributing data to users. Most of the EOS standard data products are produced at the Science Investigator-led Processing Systems (SIPs) and sent to the Data Centers for archiving. The Data Centers interoperate with each other and also with several

* This chapter was prepared as a part of the authors' employment with the U.S. Government. The opinions expressed here are those of the authors and do not necessarily reflect the official position of NASA.

international archives. A context diagram showing Data Centers, SIPs, other components of EOSDIS and external interfacing entities is presented in figure DD2. This figure shows an end-to-end view of the EOSDIS context that runs from data collection on the left, to data dissemination on the right. Data are collected from the satellite and sent to a facility for backup and initial processing. Raw data are distributed to the Data Centers and SIPs (collocated with science teams) where they are processed to higher level products. The products are archived at the Data Centers and distributed to users automatically via subscriptions or via various search and order interfaces.

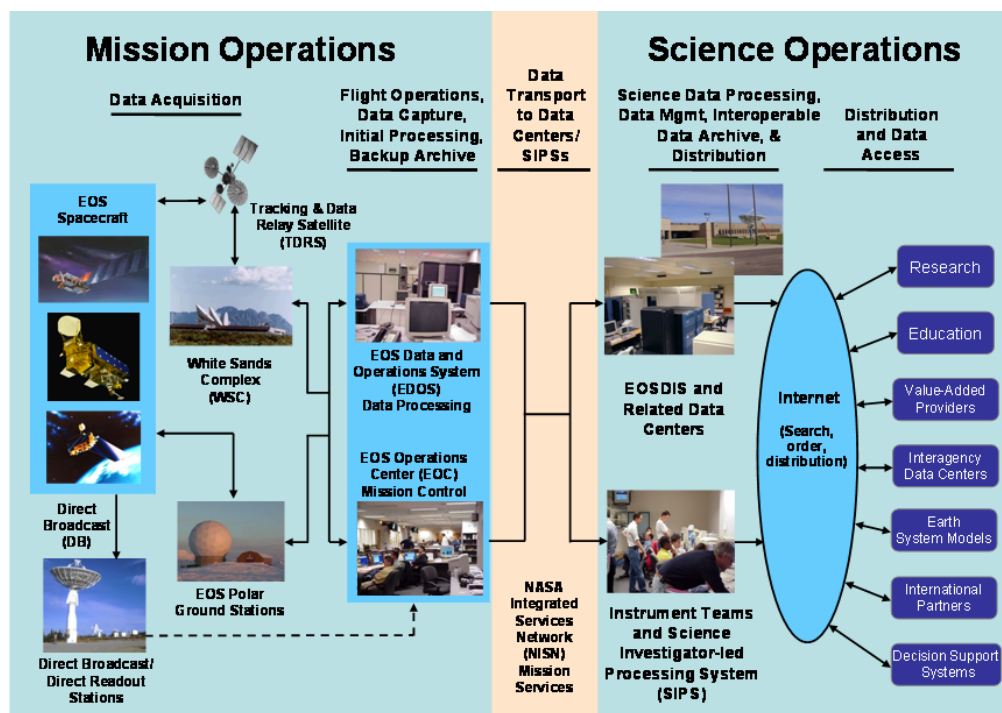


Figure DD2. EOSDIS Context

The geographic distribution of the EOSDIS Data Centers and SIPs is shown in figure DD3.

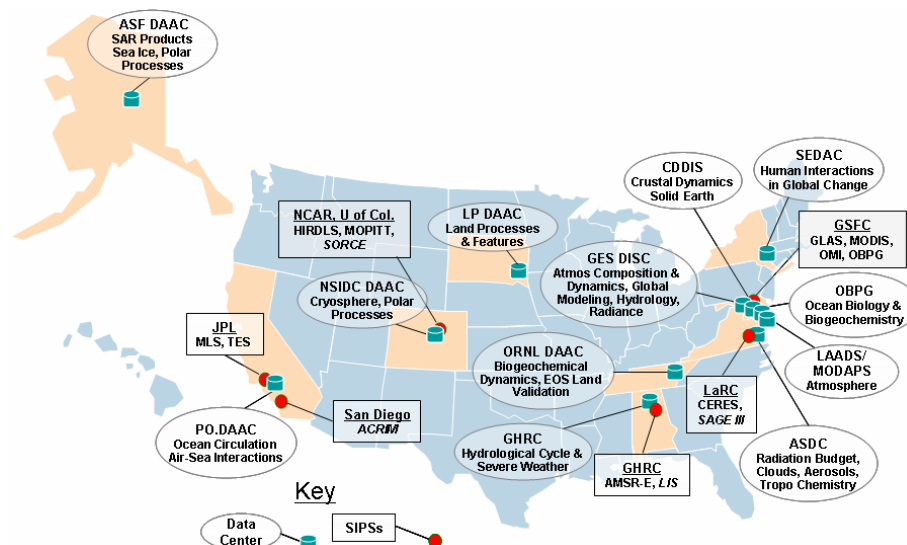


Figure DD3. Geographic Distribution of EOSDIS Data Centers and SIPs

EOSDIS provides general-purpose tools that enable location of and access to all data held in the EOSDIS system and to related Earth science data held by NASA's international partners. Such tools include the EOS Data Gateway (EDG) which is evolving into the Warehouse Inventory Search Tool (WIST), EOS Data Pools, and the EOS Clearinghouse (ECHO). EDG and WIST provide general search and access capabilities to the EOSDIS repository. Data Pools support direct, on-line access to high-priority data in individual EOSDIS archives and can be accessed either directly from the Data Pool's interface at each of the Data Centers, from EDG or WIST, or via a third-party user interface built on ECHO. ECHO is a data and service middleware tool that provides application program interfaces (APIs) upon which users can build their own machine-to-machine or human-machine search clients. In combination, EDG, WIST, Data Pools and ECHO enable flexible and efficient access to data held in the EOSDIS archives. In addition to these general interfaces, EOSDIS also supports discipline-specific access mechanisms through its Data Centers. Information about and access to the EOSDIS data access tools can be found at <http://nasadaacs.eos.nasa.gov/>.

As of September 2007, EOSDIS had over 4.9 petabytes of data in its archives. Data continue to be generated and stored at an average rate of 3.2 terabytes/day. During fiscal year 2007 (year ending September 30, 2007) EOSDIS distributed over 100 million products to end users at the rate of about 4.2 terabytes per day.

HISTORY AND EVOLUTION

Since the original formulation of EOSDIS in the late 1980s and its initial design in the early-to-mid-1990s, information technology has been revolutionized. This has led to significant changes in needs and expectations of the user community. As a result, EOSDIS has been and continues to be an evolving system in response to these changes which have driven the evolution of the basic architecture and design of the data dissemination components. Also, NASA is complementing its core infrastructure consisting of EOSDIS and mission data systems with community components selected through peer reviewed competitions. Such components include Research, Education and Applications Solution Network (REASoN) and Advancing Collaborative Connections for Earth-Sun System Science (ACCESS) projects. Thus NASA is evolving its Earth science data systems towards a more distributed, heterogeneous and on-line/easily accessible, environment. Figure DD4 illustrates the evolution of EOSDIS over its first decade of operations.

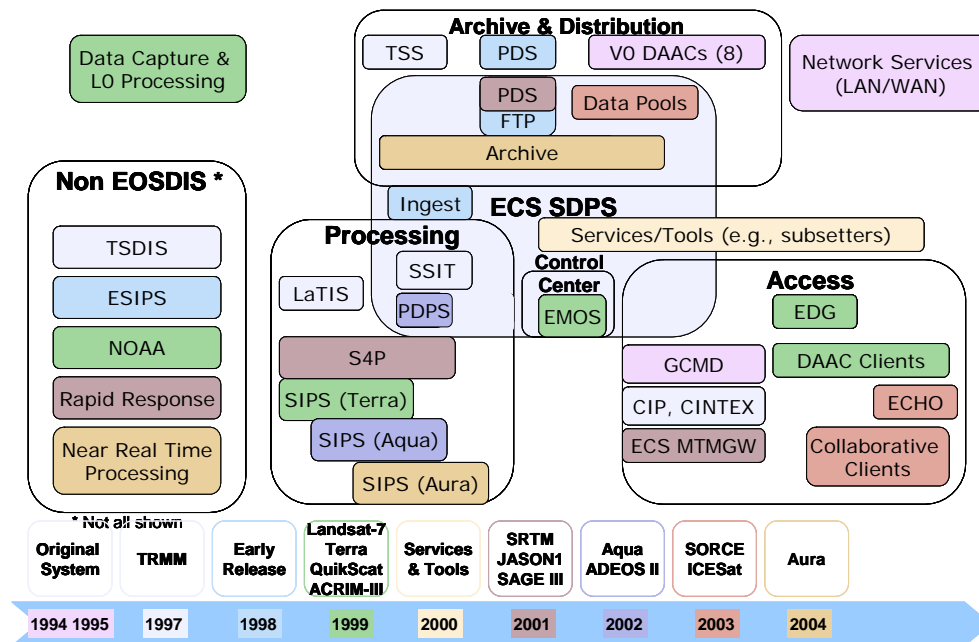


Figure DD4. Historical Evolution of EOSDIS (acronyms in Appendix A).

Version-0 IMS

EOSDIS started with 8 heritage archives that each held data for a specific Earth science discipline, and housed the science expertise to ensure proper stewardship and use of these data. The design concept was based on the philosophy that these heritage archives held the expertise and therefore the data should not be relocated to a centralized archive. The development focus was on an “inventory level” interoperability system that would support a single-point search and access interface that would distribute searches to the distributed archives and would receive and organize results sets (of specific files or “granules” meeting user-specified criteria) for viewing and ordering by the end users. The Version-0 (V0) Information Management Subsystem (IMS) was a proof-of-concept “working prototype” for interoperability among distributed archives (Ramapriyan and McConaughy, 1991). The four-year V0 development effort (1990 – 1994) was coincident with a major information technology revolution that directly impacted the technology on which this system was based. The challenges in balancing user requirements for system functionality, user expectations (driven by newly available desktop technologies), and technology maturity issues made for quite an exciting and challenging development environment. As technology evolved, the user community expected the user interface to be presented in the latest technology on their desktop. In the beginning of V0 IMS development, character-based user interfaces that ran on VT100 terminals were the standard. After a year into prototyping, X-Windows-based graphical user interfaces (GUIs) took over the market. At the operational release of the X-Windows-based V0 IMS GUI, HTML interfaces emerged and quickly took over the market. Just as users were getting used to the cumbersome click-and-wait interaction of HTML, Java emerged as a more interactive option. The V0 IMS development team could barely develop basic functionality before it needed to incorporate a new user interface technology.

The V0 infrastructure was designed to provide basic data search, browse (viewing a sample image), and order functions to users ranging from Earth scientists to the general public. This functionality was achieved by using metadata that was commonly used across all of the archives. By limiting the

set to only common attributes with consistent definitions, search criteria were limited to those that describe whole data sets rather than individual data granules (smallest inventoried unit of data in the system). However the system needed to be able to return individual granules to the end user. So, the system was architected so that the V0 IMS would distribute search messages out to the individual Data Center. The archives would execute the searches, each using their own, autonomous mechanisms, and then pass the results back to the V0 IMS client. The V0 IMS client would assemble the results sets and display them for end users in a web-based graphical user interface. Artifacts of this distributed query architecture drove a degree of complexity that sometimes resulted in confusion or frustration to the end user. This lesson learned was taken into account in developing the EOS Clearinghouse (ECHO) approach described later.

Despite the project's challenges, major successes came from the V0 IMS development effort. V0 IMS was one of the first of its kind in providing single-point search and access to distributed holdings. In addition, as a result of this effort, the V0 IMS Protocol was developed. This protocol formed the basis of future protocols used by the CEOS (Committee on Earth Observing Satellites) Interoperability Experiment (CINTEX) that connected international archives with those from EOSDIS. This international protocol became known as the CEOS Interoperability Protocol (CIP). The V0 IMS and CIP Protocols and related message passing dictionaries were fundamental inputs into the creation and evolution of current interagency and international standards and protocols.

Overview of EDG and WIST

In 1998, the Earth Science Data and Information System Project (ESDIS), which is responsible for the development and operation of EOSDIS, made the decision to use the V0 IMS as the basis for the search and order interface to EOSDIS data to support the first major milestone, the launch of Landsat-7 in April 1999. To meet the Landsat-7 launch milestone and support granule level metadata searching for Landsat-7 and the subsequent EOS missions, an extension of the V0 IMS, renamed EOS Data Gateway (EDG), was used. The initial EDG provided basic end-to-end data search and order capabilities. Over time, in order to meet specific needs of EOS instrument teams and discipline-specific science communities, new capabilities were added including some data-related services such as subsetting.

During this time, ESDIS realized that the long-term needs in EOSDIS were to support a broader set of interoperability requirements. These requirements included the ability to support search and access by alternative clients and the flexibility to support data-related services. As a result, the EOS Clearinghouse (ECHO) was initiated as an enhancement to EOSDIS. Thus the two efforts, EDG operations and ECHO development, were carried out in parallel.

As ECHO development continued, ESDIS decided that ECHO would replace the middleware portion of EDG and that the Warehouse Inventory Search Tool (WIST), a user interface similar to EDG that interacts with the ECHO middleware, would replace the EDG user interface. The resulting end user experience is similar in EDG and WIST. Therefore, they are described together below.

KEY USER INTERFACES FOR EOS MODIS AND ASTER DATA, AND PRODUCTS

EOS Data Gateway (EDG) and the Warehouse Inventory Search Tool (WIST)

Until recently, the EDG search-and-order tool (Moses, 2003) has been the primary access point to 2,100 EOSDIS and other Earth science data sets archived at the EOSDIS data centers and several international data centers. During 2007, a total of 73,383 orders were placed and 7.47 million products were ordered via EDG. As of this writing, all EDG users are being transitioned to WIST, and EDG will be retired. WIST can be accessed from <https://wist.echo.nasa.gov>. At the top of the WIST homepage, the user is presented options to login or create an account. As a registered user, users may be given access to data that is restricted to public, guest users. A registered user can also update their profile with their contact, shipping, and billing address, which can be recalled when the user orders data.

The main search interface of WIST is very similar to the EDG system behavior, which reflects architecture driven by technologies available in the early-to-mid-1990s and the requirement to meet many different users' needs. While the EDG and WIST user interfaces are complex, users have found that it is the most efficient way to find and access data and related information from EOSDIS. With the introduction of ECHO, it is expected that user communities will create their own views into ECHO which are tailored and more useful to their discipline-specific user community. ECHO and WIST provide tutorials for the users, as well as, user services contact information to which users should send questions or feedback on the system or on the data.

Searching for Data

The Primary Data Search on the WIST homepage is the default search setting and is divided into four major search criteria areas: data keywords, spatial criteria, temporal criteria and specialized search constraints (e.g., only granules with browse, day/night flag). The homepage also allows a registered user to save a search so that it can be restored during a later session, and of course, to restore it as needed. Alternative search options are available for users who know the Data Center unique identifiers for the data of interest (referred to as Data Granule ID and/or Local Granule IDs) and simply need to order them from the archives. Users can choose between these search types using the toggle options in the middle of the search screen in the Choose a Data Search Type section. All of these types are fully described in the "help for this page" link that is available at the top every page. Also, the title at the top of the search screen indicates which search type is active at a given time.

The first section of the Primary Data Search option is the data keyword selection. The default keyword selection allows users to select keyword criteria "By Discipline." However, there are important alternatives to selecting by discipline that are briefly described in the paragraphs below as they can be very useful in specifying search criteria.

The first special search feature is the ability to enter criteria into the "Text Search," if the user knows exactly what is needed. For example, if the user knows a data set short-name, or a geophysical parameter name, the user needs only to enter that term into the text box at the top right of the search page and select the "go" button to the right of the search box. The search will return a list of data sets that satisfies the term(s) entered into the text search box. The user can then scroll the list of resulting data sets and select as many as needed for further searching.

The second special search feature is the ability to be guided by listings of available content that can be selected for various data characteristics (attributes). This mode has two benefits. First, it helps users who are unfamiliar with specific data set names in EOSDIS, but do know how to characterize data of interest (e.g., Land Cover over Eastern US for January 2007). Secondly it helps data set

experts who need to query by detailed data set and product-specific characteristics. This option can be initiated by scrolling past the list of discipline toggles in the “Choose Dataset” section and selecting the “By Category/Attribute” toggle. This feature was initially developed to assist users who are unfamiliar with EOSDIS data terminology to navigate the criteria so that informative selections can be made. The extension of this capability to include very specific data set granule- and product-specific terminology might make this difficult for novices to use. Although data set experts gain the most benefit here, a novice can still benefit by confining the use of this feature to searching by data set-level attributes (i.e., characteristics that describe whole data sets). Data set-level attributes include Campaign, Data Set, Parameter, Processing Level, Instrument (sensor) and Source (platform). The novice user can select a category (e.g., Instrument), then select the specific instruments of interest (e.g., MODIS, ASTER). Based on these selections the system will only display applicable (valid) categories and characteristics in subsequent data set category lists. The user should be cautious when exploring more detailed or product-specific categories. Data set descriptions are not equally populated. For example, ASTER products do not have a “CLOUDPERCENT” in their descriptions. If a user selects the MODIS instrument and the ASTER instrument from the Instrument category, then selects to search by a “CLOUD PERCENT” range in the “Cloud Amount” keyword section, the user will not see ASTER data returned in the results sets. Similarly, if the user selects to search by Day/Night Flag from the Choose Additional Options section, the user will not see ASTER data returned in the results sets because the system will not find a match for “Day/Night Flag” for ASTER data. On the other hand, if a user selects a single instrument, or a single data set, the user can navigate other categories and content listings to view valid information, and thereby use this feature as a tool to become more knowledgeable about EOS data product metadata.

Searching by spatial, temporal and other optional criteria is described in the on-line tutorial. Once the search is constructed and initiated, WIST searches ECHO to find the data holdings that match the search parameters. Whereas the EDG search takes place at each archive site using the database search mechanism specific to that archive and may return results with different terminology standards, WIST sends its searches to the ECHO search engine where more standards can be enforced. The results will therefore be presented in a more consistent manner.

Navigating Search Results

The on-line tutorial describes a straightforward navigation through the results sets. Some unique features of navigation are described below. The first is the capability to customize the results table. With this capability, the user can set how many results items to display on a page, change the metadata columns that are displayed, change the order in which the columns are displayed, and choose different columns on which to sort at up to three levels. The second feature is the ability to save a text-only version of the results table. This allows the user to save the results set into a table for printing or importing into a spreadsheet for more flexible analysis.

When it is necessary to perform multiple searches to collect a results set, it is important use “My Folder” to save desired granule results from one session to the next. Otherwise, starting a new search clears the prior results sets and they will be lost.

Up to 20 data granules can be selected and displayed for inter-comparison of geographic footprint coverage using the “Show Map Coverage” button. This display also shows a box that represents the geographic search area that was specified in the search criteria by the user. Similarly, temporal coverage can be compared using the “Show Time Coverage” button.

Browse Imagery

Sample images or other forms of summarized content from the data granules can be displayed for “browsing” through the data prior to ordering them. In most cases browse images provide a visual display of subsampled data. They can also be sample 2-dimensional plots or tables depending on the particular product. Users can view a sample of the image using the “Image Quicklook” feature will display a JPEG version of the image within the WIST session. In addition, the user can download the original browse image onto the desktop for visualization in a tool other than WIST, although not all products provide an easily viewable browse file.

Data Access

Many of the data sets are available on-line on the Data Pools and other online archives as discussed in more detail below. Granules of such data sets are indicated on the results screen in the WIST user interface under the column “On-line Access.” They can be downloaded by the user directly by clicking on the appropriate granules listed in the results screen.

Ordering

For data that are not available for on-line access, the WIST user interfaces permit the user to move selected granules to a shopping cart, select ordering options (ftppush, ftpull or media), and place an order. The data granules are then fetched from the data center archive and staged for delivery by the user-selected mechanism. The user is informed via e-mail about the order status. It should be noted that in recent years the demand for data distribution via media has been decreasing and most users get the data electronically.

Subsetting

Using WIST, users can find granules that meet their search criteria. However, a granule may contain significantly more data than desired by the user. It may intersect with the spatial region identified by the user, but may have data that are not in the region of interest. The granule may contain many parameters, but the user may be interested in only a few. It is advantageous both to the user and to the data center providing the data to reduce the data granule to the subset of interest. Capability to do this is not uniformly available for all the data sets held by EOSDIS and is handled on a data set by data set basis. At present, options are offered by some of the Data Centers through the order options, but the actual subsetting is invoked independently of WIST for the data sets for which this capability is available. Information about the subsetting capabilities available for specific data sets can be found in the respective Data Center’s web site. Plans for the future include the ability to invoke subsetting (and other services such as reformatting and reprojection) through the ECHO middleware.

Data Pools

As Earth science applications become more interactive and/or automated, the community’s need for high-performance, direct-access systems increases. The EOS Data Pools approach makes the high-priority, current data products available for direct on-line access. [Moore and Lowe, 2002]. Several data centers are now taking advantage of this concept. In particular, the three EOSDIS Data Centers where the EOSDIS Core System (ECS) is currently operating use Data Pools that are built into the ECS. These Data Centers are: the Atmospheric Science Data Center (ASDC) Distributed Active Archive Center (DAAC) at NASA Langley Research Center (LaRC), the Land Processes (LP) DAAC, and the National Snow and Ice Data Center (NSIDC) DAAC. While they share common

software, these Data Pools are operated independently. Each Data Center determines how (and which) high-demand data are to be cached to meet its user community's needs. Users can send a request to the Data Center user services office for specific products to be placed in the Data Pool as they are inserted to the archive, tailoring the resulting content to meet the specific needs of the Data Center's user community. Once a product is in the Data Pool, it is available to all users. Generally the data flow into and out of the 'pool' on a first-in, first-out basis. The lifetime of a product in the Data Pool depends on the size of the pool itself. As of this writing the total capacity of the Data Pools at the three DAACs indicated above is about 400 Terabytes. The on-line storage capacity at all the Data Centers (including ECS and non-ECS components) is about 2 Petabytes. As the cost of disk storage decreases, more data can be made available on-line. As indicated above, EDG and WIST can be used to find data that are accessible on-line from the Data Pools as well those that are off-line and need to be staged for FTP transfer (or copying to media).

Data Pools' data access interfaces provide a simple browse and click navigation where users can make selections to 'drill down' to specific products. With each selection, the user reduces the number of granule results returned. Once the results are narrowed down to specific granules, the user can directly access the data via FTP. The user can also view the metadata and browse for the granules. The user can 'drill-down' using the following search parameters: Data Group, Data Set, Spatial, Date, Time, Cloud Cover, Day/Night Flag and Science Quality Flag. The Data Group parameter is the grouping of the data based on the instrument, mission, and major discipline. The Data Set parameter is the primary identifier for the ECS data collection. The Data Group and Data Set parameters are the only required fields. The Spatial parameter is the geographic coverage, represented by latitude/longitude points, in which the granules of interest should overlap. The spatial region is selected by drawing a rectangle or a polygon on the map applet or by specifying the latitude and longitude coordinates for a bounding rectangle in text fields. The Date parameter is the acquisition date range of the granule. This parameter is selected by clicking a month, week, or day on a calendar or by selecting a beginning date and ending date. The Time parameter is the acquisition time of day for the days within the selected data range. The time of day is selected by clicking on a particular hour increment in the table or by selecting a beginning time and ending time. The Cloud Cover parameter is a percentage indicating how much of the granule is covered by clouds. The Day/Night Flag parameter indicates whether the granule was collected during the day, night, or both. The Science Quality Flag parameter indicates the science quality (e.g. Passed, Failed, Not Investigated) of a granule measured parameter.

The granules that match the selected search parameters are displayed as individual rows in the results table. For each granule, the results table displays the granule identifier, granule size, beginning and ending date and time, and day/night flag value. Each row also has linked icons that allow the user to download the data with or without compression, access the full granule metadata record, access the browse image, and add the granule to a shopping cart. The user should add granules to the shopping cart if the user would like to have the granules processed (e.g. reformat, reproject, subset) using the HDF-EOS to GeoTIFF (HEG) conversion tool before they are delivered.

LP DAAC's Data Pool

The LP DAAC's Data Pool provides selected data products free of charge at <http://lpdaac.usgs.gov/datapool/datapool.asp>. The available geographic coverage varies by product. Geographic coverage for ASTER data products includes the United States and Territories. ASTER data have no scheduled removal from the Data Pools. Global coverage is available for MODIS

Land products. Daily MODIS products are in the pool for at least ten days after acquisition. A minimum of 12 months of acquisitions are available for other MODIS products.

NSIDC DAAC's Data Pool

The NSIDC DAAC's Data Pool provides selected data products free of charge at http://nsidc.org/data/data_pool/index.html. Global coverage is available for most MODIS Snow and Ice products. MODIS products are in the pool for about thirty days after acquisition.

EOS Clearinghouse (ECHO)

ECHO was initiated to expand the search and access options for users trying to access EOSDIS data, allowing users to develop their own tailored user interfaces and applications rather than only providing the “one-size-fits-all” solution, EDG. ECHO is middleware that facilitates the sharing of data products and data services in the Earth science community (Burnett and Wichmann, 2005; Wichmann and Pfister, 2002). ECHO provides a metadata registry and a data services registry made of EOSDIS data and contributions from the wider science community. Via these registries, ECHO enables clients and users to search, browse, place orders, and directly access on-line data across multiple resource providers. ECHO partners with data providers to ingest their metadata and provides application programming interfaces (APIs) for data providers to restrict their metadata using access control lists. ECHO offers open interfaces for clients (machine-to-machine or human-machine) to find and use these products and services. ECHO's capabilities are available to the community via a set of internet accessible APIs. ECHO provides client APIs for user registration, searching, and ordering. The Catalog Service API allows users to search for data sets and granules using Spatial (e.g. point, line, polygon, multipolygon, circle), Temporal (e.g. date range, day/night/both), Keyword (e.g. data set id, sensor name), Numeric (e.g. cloud cover percentage), and Boolean (e.g. Only data with browse data, Only data that is online) search parameters. Once the data of interest are found, the user can directly access the data if they are online or can place an order, which ECHO will broker to the appropriate data provider. ECHO also allows users to establish data subscriptions. With these APIs, clients are free to provide community-specific views or to implement new technologies for interacting with users. Clients can be in the form of a GUI that provides a discipline- or application-focused search and access service for a particular community, or in the form of a set of machine scripts driving functions within an application, modeling or decision support system. In either case, the ECHO middleware is hidden from the end user.

The ECHO website, <http://eos.nasa.gov/echo>, gives information needed for data and service providers and client developers to become partners in ECHO. It provides operational system status, information for and about data, client and service partners, ECHO development status and future plans, a mechanism for submitting comments and feedback, latest ECHO news and announcements and references including presentations from ECHO partner training sessions.

Global Change Master Directory (GCMD)

The GCMD (Olsen, 2000), <http://gcmd.nasa.gov>, holds more than 25,000 Earth science data set and service descriptions relevant to the Earth science community. Specific data set and service descriptions can be found by using the ‘drill-down’ interface made up of discipline-specific keywords. Each description includes information, such as a text summary, geographic and temporal coverage, data archive center information, and location of the data or service, which can help determine whether the data or service meets the user's needs. Users can use the GCMD to search

for information about Earth science data and services. The users will then be directed to the EDG, WIST or Data Center-specific systems to order or access EOSDIS data.

Interfaces Specializing in Land Processes Data

In addition to the general EOSDIS user interfaces (e.g. EDG, WIST) for searching and ordering or accessing data, there are several discipline-focused user interfaces offered by the Data Centers and their host organizations. This section presents three examples that are tailored to searching and obtaining land processes data. Other examples are discussed elsewhere in chapter xx of this book.

<<Provide references to other chapters.>>

United States Geological Survey: Global Visualization Viewer

The Global Visualization Viewer (GloVis) developed and operated by the USGS is an on-line, browse-based search and order tool. This tool offers access to all available browse images from EO-1, Landsat (1-5 and 7) and MRLC from the USGS inventory, and ASTER and MODIS from the LPDAAC inventory. In using GloVis, a user starts with a graphic map display of the world and can select any area of interest. GloVis displays a browse image of the selected area and the adjacent scene locations immediately. The user can pan and select any of the scenes around the area of interest. The user can also select the observation dates for the specific regions of interest. The selected images can be added to a list of items to be ordered. GloVis provides additional features such as cloud cover information, date limits, display of user-selected map layers (e.g., political boundaries, roads, railroads), and access to metadata. GloVis has an ordering interface that permits the user to specify the images to be ordered, formats in which the user would like them, and the type of media or electronic delivery desired. The ordering interface also provides the information about the price (if any) and has a mechanism to collect payments from the user.

Oak Ridge National Laboratory's Mercury System

The Mercury system, developed by the Department of Energy's Oak Ridge National Laboratory, allows users to search the ORNL DAAC and other providers' data. The Mercury system harvests metadata on a regular basis to facilitate searches and directs users to other providers' sites for access to data. The providers (e.g. Land Validation team, Safari 2000 campaign, Long Term Ecological Research Network) can select which of their metadata and data are made visible via Mercury by maintaining text files of URL lists. Mercury uses these lists to harvest the metadata on a daily basis. In order to search and order data, a user can specify a set of keywords, spatial and temporal bounds, and/or a set of sources (databases). The user is then provided with a display of the database names and the number of entries found in each of the databases matching the user's specifications. By clicking on the database name, the user can display a summary of the metadata. The metadata summary includes links to the data sets which can be downloaded either immediately or later. A shopping cart facilitates accumulating data sets to be identified and downloaded together as a group.

MODIS Search 'N Order Web Interface (SNOWI)

The MODIS Search 'N Order Web Interface (SNOWI) is a quick and simple way to search and order MODIS gridded Snow and Ice products from the NSIDC DAAC. The user specifies a few search parameters, such as data set name, temporal date range, and latitude/longitude coordinate range.

FUTURE

As discussed above, EOSDIS has been serving a broad user community by disseminating large amounts of data on a regular basis for the past several years. EOSDIS has undergone several technologies and requirements changes since it was conceptualized in the late 1980s . There have been incremental improvements in processing and performance, while new functionality has been added in user access, distribution and archive management. The underlying design, however, remained essentially the same until recently. Through a recent examination of EOSDIS' operations and lessons learned, there has been a desire to achieve significant improvements in a variety of areas. To accomplish this, in 2004 NASA established an EOSDIS Elements Evolution Study Team (EEEST) and an EOSDIS Elements Evolution Technical Team (EEETT) (Esfandiari et al, 2006). The role of the EEEST was to provide an external viewpoint and guidance. The EEEST developed a data systems' vision for the year 2015. The EEETT was also responsible for developing an approach and an implementation plan to fulfill the objectives set forth in the vision. The results of these efforts are published at <http://eosdis-evolution.gsfc.nasa.gov>. The goals established in the Vision 2015 parsed into IT systems management tenets are shown here in Table DD-1.

TABLE DD-1: EOSDIS Evolution Vision Tenets and Goals

Vision Tenet	Vision 2015 Goals
Archive Management	<ul style="list-style-type: none"> • NASA will ensure safe stewardship of the data through its lifetime. • The EOS archive holdings are regularly peer reviewed for scientific merit.
EOS Data Interoperability	<ul style="list-style-type: none"> • Multiple data and metadata streams can be seamlessly combined. • Research and value added communities use EOS data interoperably with other relevant data and systems. • Processing and data are mobile.
Future Data Access and Processing	<ul style="list-style-type: none"> • Data access latency is no longer an impediment. • Physical location of data storage is irrelevant. • Finding data is based on common search engines. • Services invoked by machine-machine interfaces. • Custom processing provides only the data needed, the way needed. • Open interfaces and best practice standard protocols universally employed.
Data Pedigree	<ul style="list-style-type: none"> • Mechanisms to collect and preserve the pedigree of derived data products are readily available.
Cost Control	<ul style="list-style-type: none"> • Data systems evolve into components that allow a fine-grained control over cost drivers.
User Community Support	<ul style="list-style-type: none"> • Expert knowledge is readily accessible to enable researchers to understand and use the data. • Community feedback directly to those responsible for a given system element.
IT Currency	<ul style="list-style-type: none"> • Access to all EOS data through services at least as rich as any contemporary science information system.

While all these goals are important in achieving the Vision, from the point of view of data dissemination systems, those relating to the tenets titled “EOS Data Interoperability” and “Future Data Access and Processing” are particularly significant. Currently, a first step in the evolution of EOSDIS elements has been defined and is being implemented. It is expected that this step will move the system in the direction articulated in the Vision. Some of the characteristics of this current work are shown in table DD-2.

Table DD-2. Characteristics of First Step of EOSDIS Evolution

Action	Expected Result
Produce some of the high-volume data products on demand	Reduction of archive sizes
Increase on-line archive capacities	Improved access to frequently used and key data products
Re-architect elements of existing systems: transition to commodity-based hardware; automate more functions; streamline functions to reduce custom code	Reduced maintenance and operations costs; improved scalability
Provide more post-processing (e.g., subsetting and reprojection) of granules on demand than currently available	Better tailored products; Increased flexibility for users; Reduction in network bandwidth requirements
Consolidate operations at any given Data Center to a single system	Reduced maintenance and operations costs
Transition some of the data system functions to science team organizations	Improved alignment with day-to-day science activities
Improve middleware services including performance and easier access; provide improved methodologies to facilitate users to become value-added data providers.	Increased support for distributed, heterogeneous provider environment

The future steps in evolution specific to EOSDIS elements will depend on the outcome of the initial step indicated above. However, we can speculate here about the future, based on the Vision, in the broader community context from the point of view of data dissemination. To help in this discussion, we first define four levels of interoperability: Directory Level, Inventory Level, Data Object Level, and Service Level.

- Directory level interoperability implies that a user will be able to find, through a directory, the location of the data or service. The directory facilitates the process by providing a link (URL) to the data/service provider (DSP). The user then obtains the data or service from the DSP. This is the type of interoperability offered by the GCMD.
- Inventory level interoperability implies that a user will be able to locate from a single location, the data granules that a set of DSPs can provide. One or more data granules can be obtained (ordered or accessed) from the DSPs. This is the type of interoperability supported by WIST. The ECHO middleware helps provide this type of interoperability by being a metadata

clearinghouse and supports development of additional specialized clients with at least the same level of interoperability.

- Data Object level interoperability implies that a user will be able to locate and access data (e.g., records) within granules located at a set of DSPs and manipulate from within his/her software. Enabling this level of interoperability involves, for example, conversions of native formats into those acceptable to the software that manipulates the data. OPeNDAP is an example of a protocol that provides this type of interoperability.
- Service level interoperability implies that a user will be able to access individual data objects within granules located at a set of DSPs, link them with appropriate services, chain the services to perform a sequence of meaningful operations, execute the software in a fully distributed environment and obtain the required results.

The present state of technology offers directory level interoperability among a large community of DSPs, granule level interoperability within smaller subsets of DSPs (e.g., Data Centers) with a mutually agreed upon set of interface protocols, and data object level interoperability within subsets of DSPs that have (as in the case of granule level interoperability) agreed upon interface protocols. Service level interoperability is not common among DSPs in the remote sensing community. The Vision calls for the interoperability to be significantly more widespread than is currently being achieved wherein a highly distributed and heterogeneous set of DSPs interoperate at the service level. The following are the main challenges to achieving this aspect of the Vision.

- For data latency not to be a concern, all data that users may want to obtain need to be on-line and easily accessible. The EOSDIS has made initial strides in making more data available on-line. As the cost of storage and computation decrease it becomes more feasible to make the data available on-line or quickly computable from on-line precursor data.
- Data needs to flow where the services are and vice versa. In the former case, bandwidth becomes a limiting factor since large data sets need to be sent from the data provider's site to the service (computation) provider's site. The cost of network bandwidth has been declining quite rapidly. Therefore, this limitation should gradually disappear. In the latter case, software compatibility and data system security are of concern since it involves execution of externally developed software on a given DSP's system.
- Agreements are needed among participating DSPs regarding data formats, naming conventions and gridding schemes. Especially if the data from multiple sources are to be fed into models, applications and decision support systems, the compatibility of data formats, naming conventions and gridding schemes is critical. This requires community consensus, and open sharing of interfaces and protocols.

Progress is being made in the community (Earth sciences, other scientific disciplines and information technology) to address these challenges. For example, The Earth Science Information Partners (ESIP) Federation (with its Foundation for Earth Science - <http://www.esipfed.org/foundation/index.html>) is developing a portal to consolidate the information resources from the participating organizations and to provide basic visualization tools. The ESIP Federation consists of over 110 member organizations among which are the NASA-funded Data Centers and REASoN Projects, as well as NOAA data centers, universities, non-profit organizations, non-governmental organizations, and commercial entities. The portal, called Earth Information Exchange, is being architected as a gateway for viewing and accessing Earth science data, applications and supporting documentation. The Global Grid Forum (<http://www.gridforum.org/>), an international community representing over 400 organizations, is

leading global standardization for grid computing in various scientific and technical areas. The Global Alliance (<http://www.globus.org/>) is actively developing technologies needed for “grid” that enables sharing of computational capacity, databases and other on-line tools securely among multiple organizations while preserving local autonomy. The Federal Geographic Data Committee’s (FGDC) Geospatial Applications and Interoperability Working Group is promoting the use of georeferenced information from multiple sources over the Internet. The geospatial interoperability is based on voluntary consensus standards governing geospatial concepts and their inclusion in communication protocols, software interfaces, and data formats (<http://gai.fgdc.gov/girm/v1.1>). In addition, NASA has established a set of four Earth Science Data System Working Groups addressing technology infusion, software reuse, standards processes and metrics. These working groups have been active since 2004 and are an important mechanism for providing community best practices and guidance to NASA.

In 2005, 16 proposals were selected for NASA’s ACCESS (Advancing Collaborative Connections for Earth System Science) Program. These awards focus on improving NASA’s Earth science data systems and are also aids in helping NASA evolve EOSDIS. The overall goals of these first ACCESS projects are “... to enhance existing or create new tools and services that support NASA’s evolution to Earth-Sun System Division science measurement processing systems and support NASA’s Science Focus Area (SFA) community data system needs.” Abstracts for all these 2005 ACCESS awards can be found at http://nspires.nasaprs.com/external/viewrepositorydocument/29166/NNH05ZDA001_ACCESS_SUMMARIES.pdf. Many of these projects are applying web services architecture and open Geographical Information System (GIS) standards, specifically those from the Open Geospatial Consortium (OGC). The use of existing information technology capabilities, standards and protocols is essential for these projects to rapidly enable more robust tools and services in support of NASA’s science user communities. ACCESS projects of interest to the land remote sensing community are described below:

- R. Barry (*University of Colorado*) - *Discovery, Access, and Delivery of Data for the International Polar Year (IPY)*: Adopt principles of Web Services architecture to develop and implement a portal for cryospheric data and related information.
- A. Bingham (*Jet Propulsion Laboratory*) - *Earth Science Datacasting - Informed Data Pull and Visualization*: Based on the popular concept of podcasting, which gives listeners the capability to download only those mp3 files that match their preference, develop “earth science datacasting” to give users control to download only the Earth science data files that are required for a particular application.
- Y. Bock (*Scripps Institution of Oceanography, UCSD*) - *Modeling and On-the-fly Solutions in Solid Earth Science*: Develop a unified, on-the-fly, Web Services-based observation/analysis/modeling environment for geophysical modeling and natural hazards research, a plug-in service for early warning systems and transfer of rapid information to civilian decision makers and the media, and an educational tool.
- J. Masek (*Goddard Space Flight Center*) - *Building a Community Land Cover Change Processing System: Create infrastructure for a distributed Land Cover Change Community-based Processing and Analysis System (LC-ComPS)*. The LC-ComPS environment is envisioned as a distributed network of processing centers, linked with data archives via Data Grid technology, to allow regional and continental-scale analysis land cover at high resolution.
- K. McDonald (*Goddard Space Flight Center*) - *The Development and Deployment of a Coordinated Enhanced Observing Period (CEOP) Satellite Data Server*: Interconnect two existing technologies, the OGC

Web Coverage Server and the OPeNDAP. Utilize geospatial processing capabilities of the OGC Web Coverage Server and provide transparent data access to OPeNDAP-enabled science data applications and analysis clients used by most of CEOP scientists.

- *J. Morisette (Goddard Space Flight Center) - Improving Access to Land and Atmosphere science products from Earth Observing Satellites- Helping NACP Investigators Better Utilize MODIS Data Products:* Streamline access to MODIS data products, reduce data volume by providing only those data required by the user, and improve the utility of data products. Provide North American Carbon Program (NACP) investigators with custom preprocessing of MODIS data that will allow for direct ingest into an investigators modeling framework.

The above are a few examples of progress being made both by NASA and its affiliated user communities that are expected to have guiding influence on the future of Earth science data dissemination systems. As the history of information technology over the past decade has shown, it is quite difficult, if not impossible, to predict accurately how data systems will evolve over the next ten years. However, Vision 2015 discussed above provides the community expectations and NASA's initial steps and external developments support progress towards the Vision.

Acknowledgement

Figure DD1 is adopted from the EOS Reference Handbook, NASA, 1993, page 18, modified and updated. Figure DD2 has evolved over time from several presentations by managers at the ESDIS Project, NASA GSFC. Figure DD4 is adopted from a presentation titled "EOSDIS Elements Evolution Tiger Team Progress Report", January 20, 2004, by Glenn Iona, NASA GSFC. The authors appreciate the detailed review and comments by Frances Lindsay and Martha Maiden (NASA Headquarters) on a draft version of this chapter. They also appreciate the encouragement and support of Mary Ann Esfandiari (NASA Goddard Space Flight Center).

References:

1. Burnett, M., K. Wichmann, "ECHO - Leveraging Web Service Technologies to Support a Net-centric Science Enterprise", IGARSS 2005, Seoul, South Korea, July 25-29, 2005.
2. Esfandiari M. A., H. K. Ramapriyan, J. Behnke and E. Sofinowski, "Evolution of the Earth Observing System (EOS) Data and Information System (EOSDIS)", IGARSS 2006, Denver, CO, July 31-August 4, 2006.
3. Moore, J. M. and D. Lowe, "Providing rapid access to EOS data via Data Pools", SPIE Annual Meeting, Earth Observing Systems VII, July 7-10, 2002
4. Moses, J.F., "EOS Data Gateway: The Latest in Data and Data Access" at the American Society for Photogrammetry and Remote Sensing (ASPRS) 2003 Annual Conference, Anchorage AK, May 5-9, 2003.
5. Olsen, L.M., "Discovering and Using Global Databases," in Global Environmental Databases: Present Situation; Future Directions, Tateishi, R and Hastings, D. (Editors), International Society for Photogrammetry and Remote Sensing (ISPRS), 2000.
6. Ramapriyan, H. K. and G. R. McConaughy, "Version 0 EOSDIS - An Overview," Technical Papers, ACSM-ASPRS Annual Convention, Vol. 3, pp 352-362, March 1991.
7. Wichmann, K., R. Pfister, "ECHO - A Message-Based Framework for Metadata and Service Management", Earth Science Technology Conference, Pasadena, CA, June 11-13, 2002.

Appendix A: Acronym List.

API - Application Programming Interface
CEOS – Committee on Earth Observing Satellites
CINTEX – CEOS interoperability Experiment
CIP – CEOS Interoperability Protocol
DAACs – Distributed Active Archive Systems
ECHO – EOS Clearing House
ECS – EOSDIS Core System
EDG – EOS Data Gateway
EDGRS – Earth Data Gathering and Reporting System
EMSn – EOS Mission Support Network
EO-1 – Earth Observing -1
EOS – Earth Observing System
EOSDIS – EOS Data and Information System
ESDIS – Earth Science Data and Information System
ESSn – EOS Science Support Network
GSFC – Goddard Space Flight Center
GCMD – Global Change Master Directory
LaTIS – LaRC TRMM Information System
LTA – Long Term Archive
MBS – Measurement-based Systems
MODAPS – MODIS Adaptive Processing System
MRLC - Multi-resolution Land Characteristics
NOAA – National Oceanic And Atmospheric Administration
OCDPS – Ocean Color Data Processing System
OPeNDAP - Open-source Project for a Network Data Access Protocol
SDPS - Science Data Processing Segment
SIPS – Science Investigator-Led Processing System
SPSO – Science Processing Support Office
TSS – TRMM Support System
USGS – United States Geological Survey